

Machine Learning 1

Lecture 5.3 - Supervised Learning
Classification - Decision Regions

Erik Bekkers

(Bishop 1.5, 4.1)



Classification through decision regions

▶ Input: $\mathbf{x} = (x_1, \dots, x_D)^T$

▶ Target: $t \in \{C_1, C_2, \dots, C_k\}$

▶ 2-class targets: $t = C_1, t = C_2 \Leftrightarrow t = 0, t = 1$

▶ Multi-class targets e.g. $k=5, t = C_3 \Leftrightarrow$ $\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$
one-hot encoding of class C_3

Strategy:

▶ Divide input space \mathbb{R}^D into K decision regions.

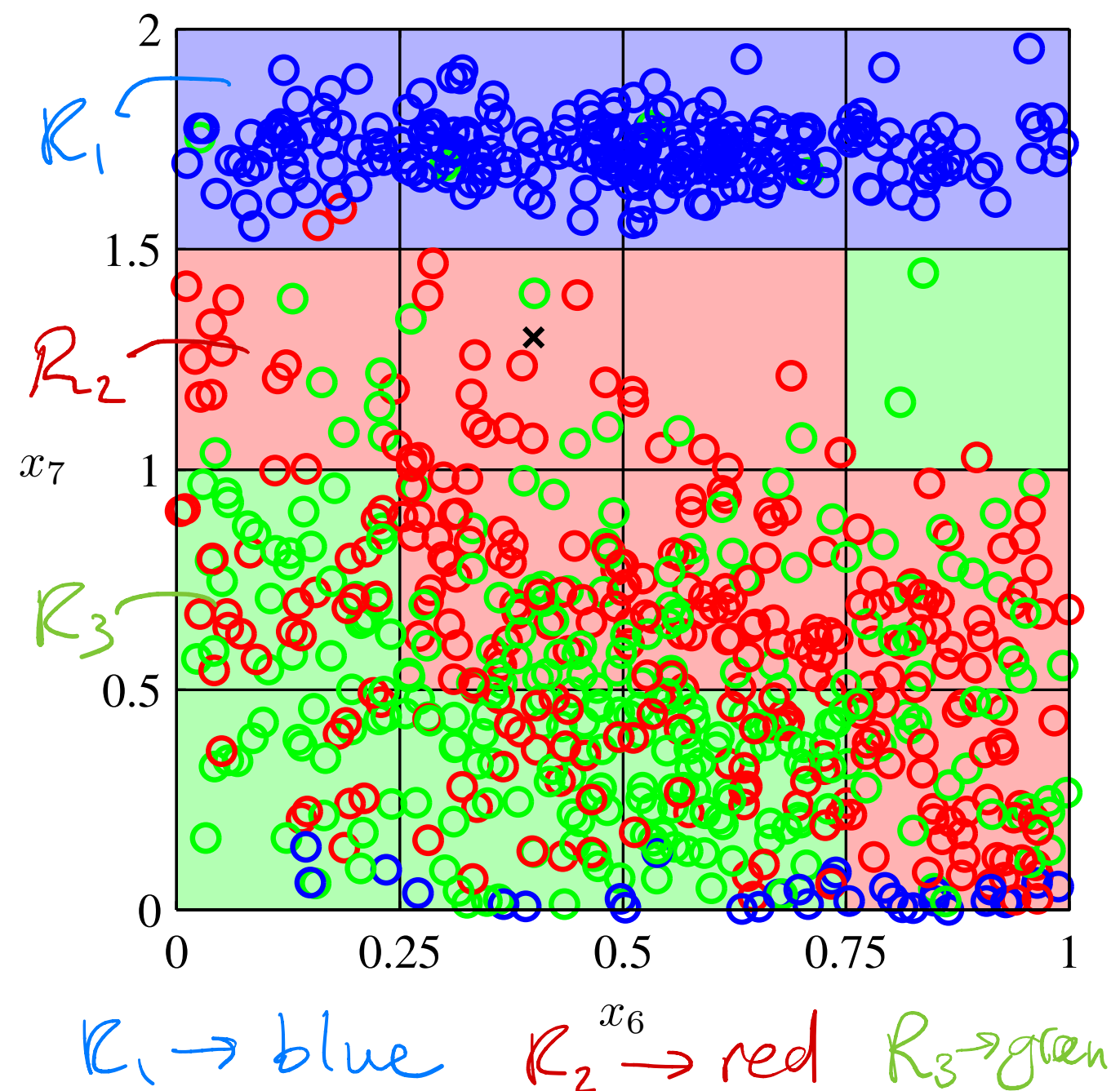
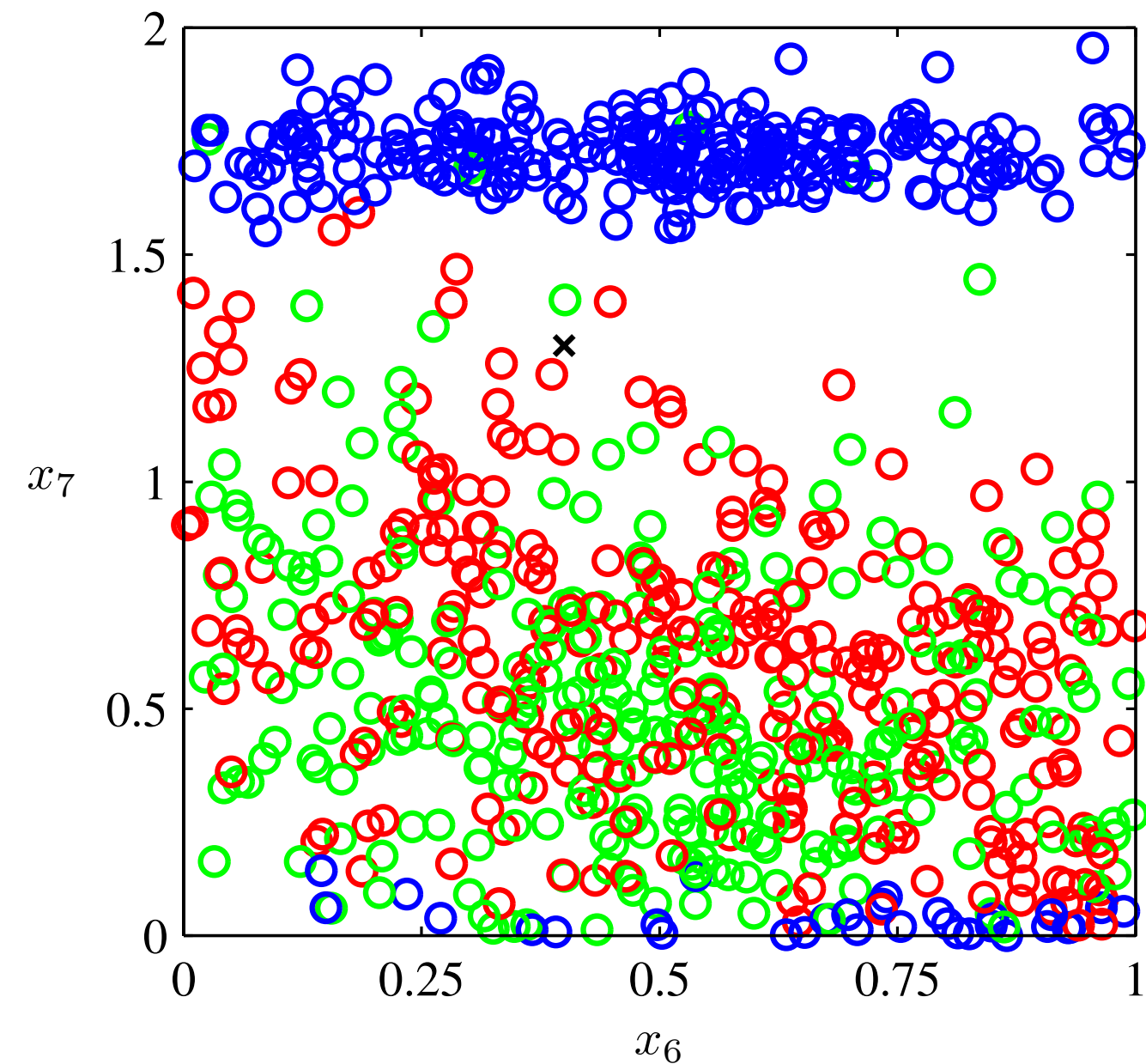
R_k

▶ Assign each decision region to a class

\downarrow
 C_k

▶ Boundaries of decision regions are called *decision boundaries/surfaces*.

Classification through Decision Regions



Figures: 3 class problem with decision boundaries. (Bishop 1.19 & 1.20)

$K=3$

Linear Classification

- ▶ **Linear Classification:** consider only *linear* decision boundaries
- ▶ For D - dimensional input space: $\underline{x} \in \mathbb{R}^D$
decision surface is a $D-1$ dimensional hyperplane
- ▶ Datasets whose classes can be separated *exactly* by linear decision surfaces are called *linearly separable*

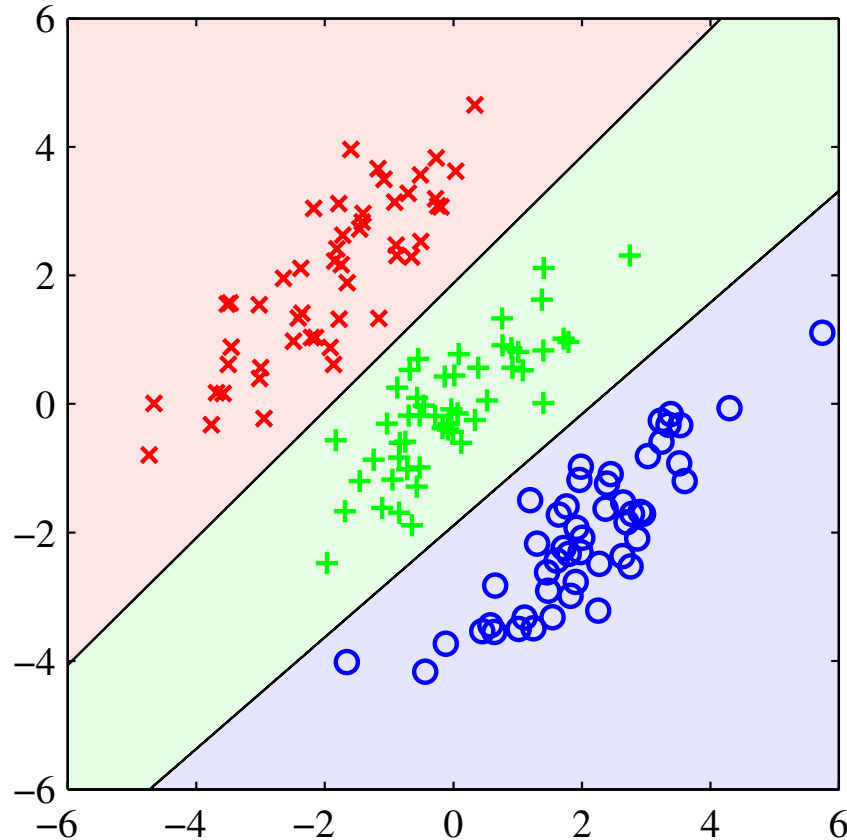


Figure: Linearly separable dataset (Bishop 4.5)

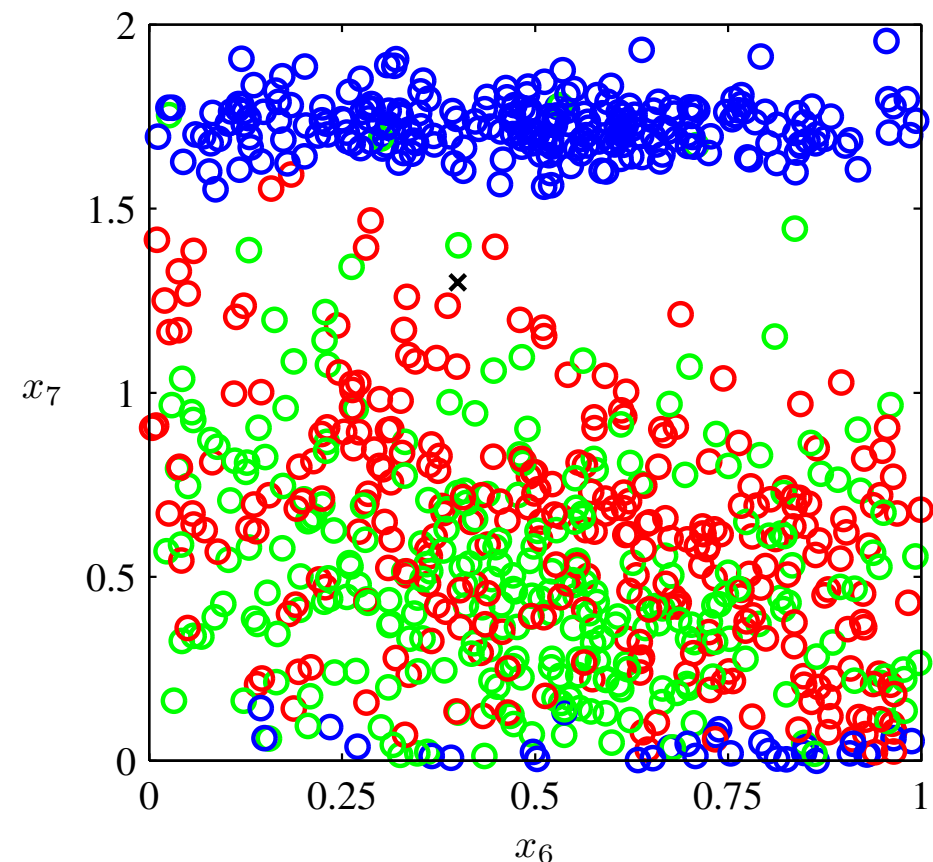


Figure: Not linearly separable dataset (Bishop 1.19)

Multiple Classes ($K > 2$)

$t = C_1 \text{ \& } C_2 ??$

- ▶ $K=2$ classes:

- ▶ 1 classifier determines

$t = C_1, t = C_2$

- ▶ Multiple classes: $K > 2$

- ▶ $K-1$ classifiers: $t = C_k, t \neq C_k$

- ▶ One-versus-the-rest

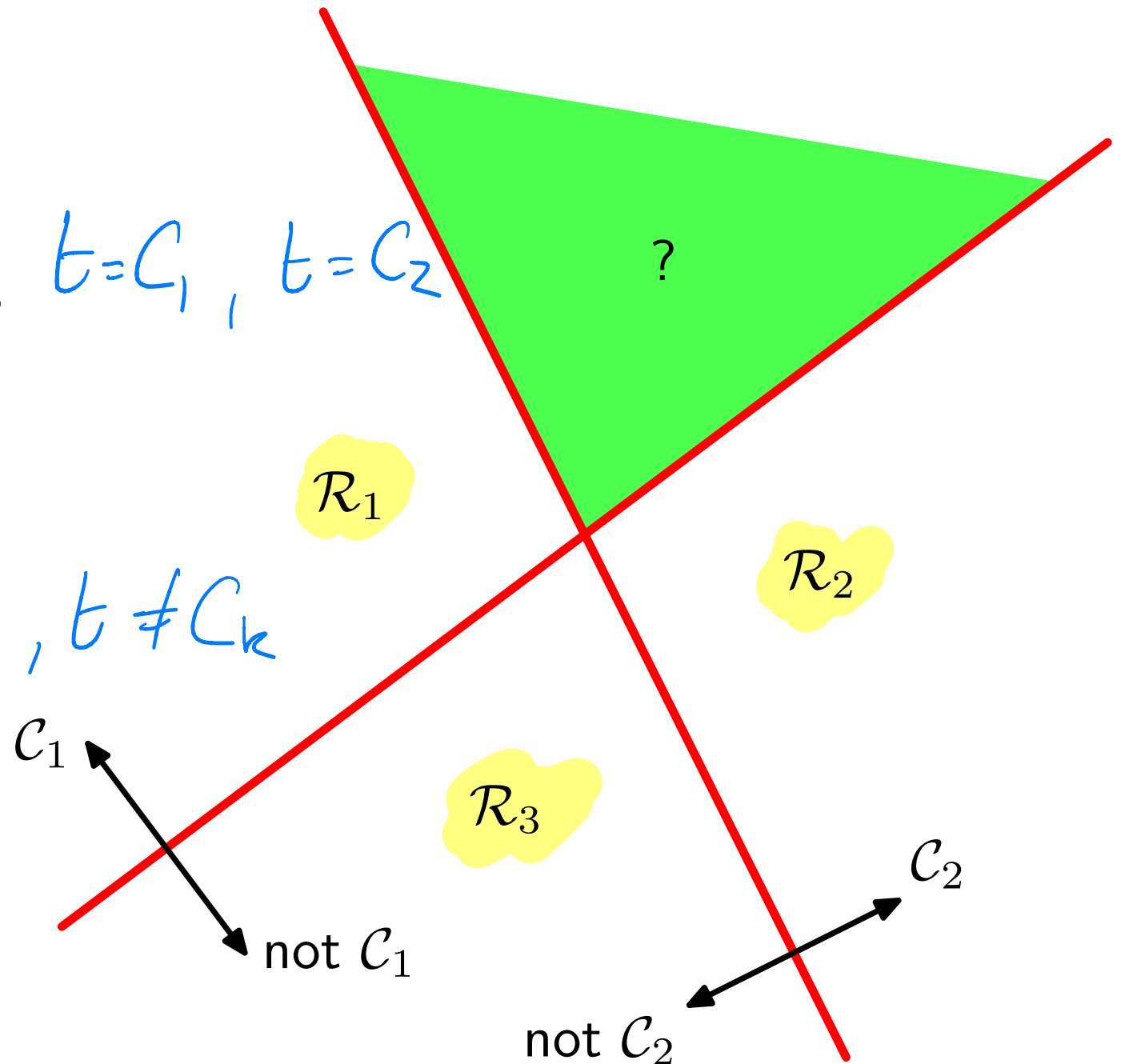
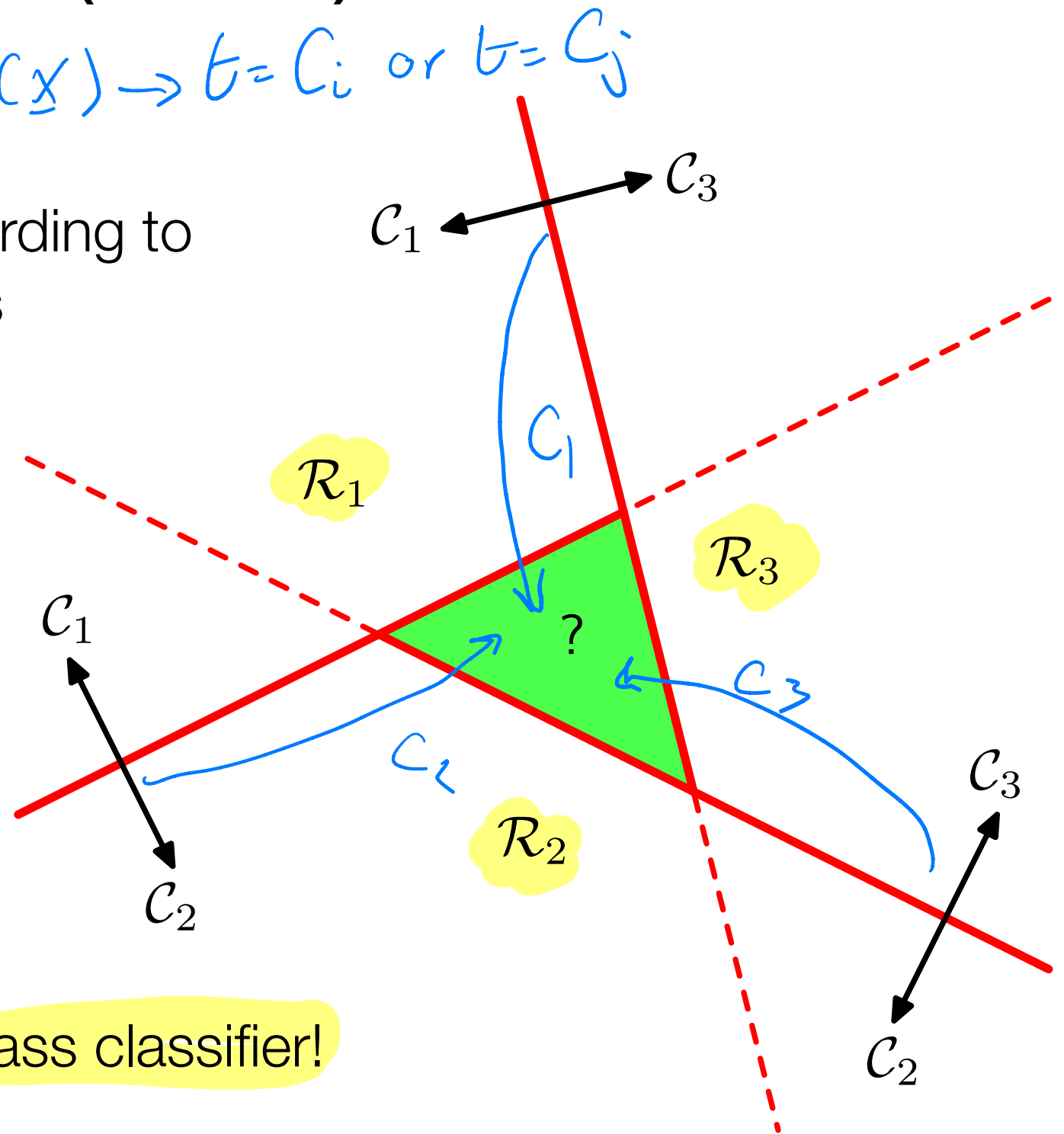


Figure: one-versus-the-rest classifiers (Bishop 4.2)

Multiple Classes ($K > 2$)

- ▶ $K(K-1)/2$ classifiers: $\psi_{ij}(x) \rightarrow t = C_i \text{ or } t = C_j$
- ▶ Points are classified according to majority vote of classifiers
- ▶ one-versus-one



- ▶ **Solution:** Make one K-class classifier!
(See later)

Figure: one-versus-one classifiers (Bishop 4.2)